

*Proteomic and peptidomic UHPLC-ESI  
MS/MS analysis of cocoa beans  
fermented using the Styrofoam-box  
method*

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Scollo, E., Neville, D. C. A., Oruna-Concha, M. J., Trotin, M., Umaharan, P., Sukha, D., Kalloo, R. and Cramer, R. (2020) Proteomic and peptidomic UHPLC-ESI MS/MS analysis of cocoa beans fermented using the Styrofoam-box method. Food Chemistry, 316. 126350. ISSN 0308-8146 doi: <https://doi.org/10.1016/j.foodchem.2020.126350> Available at <https://centaur.reading.ac.uk/88891/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.foodchem.2020.126350>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

**Proteomic and peptidomic UHPLC-ESI MS/MS analysis of cocoa beans  
fermented using the Styrofoam box method**

**Emanuele Scollo<sup>1,2</sup>, David C. A. Neville<sup>2</sup>, M. Jose Oruna-Concha<sup>3</sup>, Martine  
Trotin<sup>2</sup>, Pathmanathan Umaharan<sup>4</sup>, Darin Sukha<sup>4</sup>, Rena Kalloo<sup>4</sup> and Rainer  
Cramer<sup>1\*</sup>**

<sup>1</sup> Department of Chemistry, University of Reading, Reading RG6 6AD, UK

<sup>2</sup> Mondelēz International, Reading Science Centre, Reading RG6 6LA, UK

<sup>3</sup> Department of Food and Nutritional Sciences, University of Reading, Reading RG6 6AP,  
UK

<sup>4</sup> Cocoa Research Centre, University of West Indies, St Augustine, Trinidad and Tobago

\*Address correspondence to:

Prof Rainer Cramer, Department of Chemistry, University of Reading, Whiteknights, Reading  
RG6 6AD, UK.

Tel.: +44-118-378-4550; e-mail: r.k.cramer@rdg.ac.uk

**Keywords:** *Theobroma cacao*, cocoa beans, plant proteomics, fermentation, cocoa  
bean peptides

## **Abstract**

This work characterises the peptide and protein profiles of *Theobroma cacao* beans of the genotype IMC 67 at different fermentation stages, using the Styrofoam-box fermentation method and employing UHPLC-ESI MS/MS for the analysis of peptides and proteins extracted from the beans. A total of 1058 endogenous peptides were identified and quantified over four fermentation time points. The majority of these peptides were formed after 2 and 4 days of fermentation, and originated predominantly from the proteolysis of two storage proteins - vicilin and a 21 kDa albumin. The changes in the peptide profile over fermentation were subsequently evaluated, and potential markers for assessing the degree of fermentation were identified. In particular, changes of the relative abundance of the major cocoa proteins detected can be proposed as potential markers for the fermentation stage. Furthermore, PCA of both the peptidomic and proteomic data has allowed differentiation of beans at different fermentation stages.

## **1. Introduction**

Fermentation is an essential process for the development of cocoa flavour, and it is usually undertaken locally, on the farms, shortly after harvesting. The cocoa pods are typically opened with knives or machetes, the beans are removed and piled up into heaps or placed in wooden boxes, with a total mass of up to 1.5 tonnes (Afoakwa, Paterson, Fowler, & Ryan, 2008). Fermentation proceeds naturally, as the pulp surrounding the beans is contaminated with microflora originating from the surrounding environment. Fermentation can last 4-7 days, depending on the cocoa varieties and local practice. The cocoa mass is turned every 2-3 days to allow aeration and favour the growth of aerobic microorganisms. At the end of fermentation, the moisture content of the beans is reduced from an initial value of around 20-30% to less than 8% by drying (Afoakwa, Paterson, Fowler, & Ryan, 2008). This drying process can be done naturally in the sun, with regular turning, or with artificial dryers in closed rooms with temperatures not exceeding 60° C.

The pulp surrounding the beans is a rich medium for microbial growth and therefore an ideal fermentation substrate. Yeasts are the main microorganisms present in the early stage of fermentation, and their growth is favoured by the low pH of the pulp and low oxygen level (Schwan & Wheals, 2004). Yeast cells generate ethanol and hydrolyse pectin in the pulp walls with the release of juice and formation of void spaces into which air percolates. Lactic acid bacteria subsequently replace the yeast cells for the next fermentation stage, as they thrive in the presence of ethanol and anaerobic conditions. These bacteria metabolise glucose, with the release of lactic acid, ethanol, acetic acid, glycerol, mannitol, and CO<sub>2</sub> (Schwan & Wheals, 2004). As the oxygen level increases due

to disappearance of the pulp, and the temperature rises above 37° C, acetic acid bacteria become the main microorganisms (Schwan & Wheals, 2004). These bacteria oxidise ethanol to acetic acid, and further oxidise acetic acid to CO<sub>2</sub> and water. The exothermic reactions of acetic acid bacteria cause the temperature of the fermenting mass to increase to 50° C or more.

In addition, the penetration of acetic acid into the cotyledon activates endogenous proteases that degrade the cocoa storage proteins (J. Voigt, Biehl, Heinrichs, Kamaruddin, Marsoner, & Hugi, 1994). The released free amino acids and peptides can react with sugars through Maillard reactions during roasting, forming volatile compounds that contribute to the cocoa flavour. It is known that globulins are extensively degraded during fermentation, while the degradation of albumins is less pronounced during this process, as these proteins act as protease inhibitors (N. Kumari, Kofi, Grimbs, D'Souza, Kuhnert, Vrancken, et al., 2016; Lerceteau, Rogers, Petiard, & Crouzillat, 1999; J. Voigt, Biehl, & Wazir, 1993). However, there is some discrepancy as to the extent of albumin degradation during fermentation reported in the literature (N. Kumari, et al., 2016; Lerceteau, Rogers, Petiard, & Crouzillat, 1999; J. Voigt, Biehl, & Wazir, 1993).

Cocoa aroma can be produced from the roasting of peptide fractions, extracted from fermented cocoa beans, in the presence of sugars and deodorised butter (J. Voigt, Biehl, Heinrichs, Kamaruddin, Marsoner, & Hugi, 1994). This confirms that the peptide extracts contain aroma precursors. MALDI-TOF MS and LC-ESI MS analysis of the fractions producing cocoa aroma have revealed the presence of several peptides, whose amino acid sequences were linked to cocoa globulins (J. Voigt, Janek, Textoris-Taube, Niewianda, & Wostemeyer, 2016). It was also shown that the pH can significantly affect the type of peptides released from cocoa vicilin, as proteolysis carried out at pH 5.2 produced longer peptide sequences compared to those observed at pH 4.8 (Juergen Voigt, Textoris-Taube, & Woestemeyer, 2018).

Analysis of free (endogenous) peptides by LC-ESI MS/MS of cocoa bean samples, at different fermentation stages, have revealed that peptides generated from the degradation of cocoa vicilin during fermentation are localised in different regions of the amino acid sequence of this protein (N. Kumari, et al., 2016). The majority of the peptides observed were released during early stages of fermentation, and proteolytic activity could be observed up to 72 h from the start of fermentation (N. Kumari, et al., 2016).

Peptides formed by spontaneous fermentation and protein degradation, were extensively characterised by Souza *et al.* (D'Souza, Grimbs, Grimbs, Behrends, & Corno, 2018) employing LC-ESI MS/MS. Their results showed that during the early stage of fermentation longer peptides were predominantly released, and subsequently degraded to shorter peptides as the fermentation progressed. The identified peptides could be linked to the

action of both endo- and exo-peptidases, degrading mostly albumin and vicilin at both protein termini. A similar methodology was used to assess differences in peptide profiles between 25 samples from different geographic origins and various degrees of fermentation (Neha Kumari, Grimbs, D'Souza, Verma, Corno, Kuhnert, et al., 2018). The authors stated that the number of identified peptides was correlated to the fermentation stages, as poorly or non-fermented beans showed a lower number of peptides compared to fully fermented beans. The degree of fermentation was highly correlated to the main differences in the peptide profiles of the samples analysed, while no significant differences were found when taking into account the geographic origin only (Neha Kumari, et al., 2018).

Laboratory-based fermentation of cocoa beans, carried out in sterile glass bottles, allowed the identification of 449 peptides by LC-MS, ranging from 4 to 23 amino acid residues in length (Warren A., Bottcher, Asskamp, Bergounhou, Kumari, Ho, et al., 2019). Nine peptides, derived from the proteolysis of cocoa vicilin and formed only during the late fermentation stages, were significantly less abundant after roasting. According to the authors, this loss suggests that these peptides may be responsible for the generation of cocoa aroma (Warren A., et al., 2019).

The research presented here is based on the evaluation of the proteomic and peptidomic profiles of fermented cocoa beans using the optimized small-scale Styrofoam box fermentation method and UHPLC-ESI MS/MS for both bottom-up label-free proteomic analysis and the analysis of endogenously fragmented peptides. The Styrofoam box fermentation method was developed at the Cocoa Research Centre of the University of the West Indies, and demonstrated that similar flavour profiles can be obtained when compared to the traditional wooden box fermentation method. The advantage of fermentation in Styrofoam boxes is the much smaller quantities of cocoa beans that are required compared to the traditional wooden box fermentation. Another advantage is the overall better control and reproducibility of the whole process and its environment. The aim of this work was to evaluate whether markers for assessing the degree of cocoa bean fermentation can be selected from beans fermented using the Styrofoam box method.

## **2. Materials and methods**

### **2.1. Chemicals**

Petroleum ether 40-60 and polyvinylpyrrolidone (PVPP) were purchased from Fisher Scientific, Loughborough, UK. Urea, thiourea, dithiothreitol (DTT), water MS grade, acetonitrile MS grade, methanol, formic acid MS grade, acetone and sodium ascorbate were purchased from Sigma-Aldrich, Gillingham, UK. Trypsin was obtained from Promega, Southampton, UK.

## **2.2. Plant materials and fermentation protocol**

For all analyses, the *Theobroma cacao* genotype IMC 67 was used. Cocoa beans were harvested at the Cocoa Research Centre of the University of West Indies, St. Augustine, Trinidad. Approximately 3 hours after harvest, the pods were cut with a knife, and a total of ~3 kg of beans was placed into a Styrofoam box, which was covered with banana leaves so that fermentation could start naturally. The bean mass was turned every 2 days, and allowed to ferment for a total of 6 days. Approximately 50 g of beans were sampled at day 0, day 2, day 4, and day 6 (end of fermentation), all at the same time between 10:30 and 11:00 am local time. The temperature and pH of the beans were monitored during fermentation, and these parameters, together with a visual inspection of the beans, were used to assess the end of fermentation. Fermentation was stopped when the cocoa mass reached a pH of 3.58 and a temperature of 44° C. The aroma of the fermented mass at the end of fermentation was of a moderate acidic type.

Upon sampling, the beans were stored at -20° C with the testa removed and subsequently freeze-dried for 24 hours. The dried beans were stored at -20° C, and air-freighted in less than 96 hours to the University of Reading, UK. Upon receipt, the beans were stored at -20° C prior to analysis.

## **2.3. Defatting**

Defatting of the beans was performed following a previously published procedure (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019). Fat was extracted from ground snap-frozen cocoa beans with petroleum ether (three times) by centrifugation and discarding the supernatant. For protein analysis approximately 160 mg of ground cocoa bean powder was used, while 200 mg of bean material was defatted for peptide analysis.

## **2.4. Removal of polyphenols**

Removal of polyphenols was carried out following the protocol published by Scollo *et al.* (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019). Briefly, polyphenols were extracted from the defatted bean material three times using 3.5 ml of a wash solution (5 mM sodium ascorbate in 80% acetone / 20% water). A final extraction was then carried out with acetone only to remove residual water. The samples were subsequently dried under a stream of nitrogen to obtain acetone-dried powder (ACDP). This procedure was only carried out for protein extraction.

## **2.5. Protein extraction and Bradford assay**

The ACDP was placed on a vertical shaker for 1 hour and proteins were extracted using 3.5 ml of a solubilisation solution (7 M urea, 2 M thiourea, and 20 mM DTT in water), following a previously published protocol (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019). The supernatant obtained after the extraction was stored at -80° C prior to analysis. The amount of total protein in each sample solution was measured with the

Bradford assay (Bradford, 1976), using bovine serum albumin (BSA) as reference standard.

## **2.6. Peptide extraction**

Approximately 25-30 mg of PVPP were added to the defatted cocoa beans, and 4 ml of a 0.5% trifluoroacetic acid (TFA) in water:methanol (80:20; v:v) solution were subsequently added. The samples were vortexed for 1 minute and extracted for 1 hour at room temperature in a vertical shaker at 700 rpm. The extracted peptide solution was centrifuged at 3100 g for 10 minutes at 20° C. The supernatant was removed and stored at -80° C prior to analysis.

## **2.7. Trypsin digestion**

Digestions of the protein extracts were carried out following a procedure published by Scollo *et al.* (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019). The protein extracts were spiked with BSA, the pH adjusted to approximately 8.0 with the addition of ammonium bicarbonate, and incubated at 37° C after the addition of dithiothreitol (DTT) to a final concentration of 10 mM. The solutions were subsequently alkylated by the addition of an aqueous 200-mM iodoacetamide (IAA) solution to have a final concentration of IAA of 20 mM, and incubated at room temperature for 30 minutes in the dark. The samples were diluted with a 2-M urea solution to obtain a final urea concentration of 0.6-0.7 M. The resulting solutions were digested with trypsin at a trypsin-to-protein ratio of 1:50 for approximately 16 hours at 37°C. The digestions were stopped by adding TFA.

## **2.8. Desalting of tryptic digests and endogenous peptide extracts**

Based on a previously published methodology SOLAµ HRP 96-well plate 2 mg sorbent mass SPE cartridges (Thermo Scientific, Waltham, MA USA) were used to desalt tryptic peptides (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019). The resulting solutions were stored at -80° C prior to UHPLC-ESI MS/MS analysis. A similar procedure was followed for desalting endogenous peptide solutions. In this case an aliquot of 0.4 ml of each peptide extract solution was diluted to a final volume of 2 ml with 0.5% aqueous TFA. The same diluent was used for equilibration and washing of the SPE cartridges prior to and after sample loading. Peptides retained on the cartridges were eluted with 2x 25 µl of 0.1% TFA in acetonitrile:water 75:25 (v:v). The SPE eluates were dried in a SpeedVac and reconstituted in 0.1 ml of 0.1% TFA prior to UHPLC-ESI MS/MS analysis.

## **2.9. UHPLC-ESI MS/MS analysis of tryptic digests and endogenous peptides**

The UHPLC-ESI MS/MS analysis of the desalted tryptic digests and endogenous peptides was carried out on a Dionex Ultimate 3000 (Thermo Scientific) UHPLC system coupled to an Orbitrap Q Exactive (Thermo Scientific) mass spectrometer. The chromatographic separation was performed on an Acquity Peptide CSH C18, 150 mm × 0.1 mm ID, 1.7 µm particle size, analytical column (Waters, Elstree, UK), using a gradient with 0.1 % (v:v)



formic acid in water as mobile phase A and 0.1 % (v:v) formic acid in acetonitrile as mobile phase B. The injection volume was 15 µl for the proteomic analysis and 10 µl for the peptidomic analysis. Further details of the chromatographic conditions and MS parameters for this analysis can be found in a previously published article (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019).

## **2.10. Data analysis**

All raw MS/MS spectra files were processed using Proteome Discoverer software (Version 2.1). The obtained peak lists were loaded on the Mascot server (Version 2.6.1), and searches were carried out against the Cacao Matina 1-6 Genome v1.1 *Theobroma cacao* database ([http://www.cacaogenomedb.org/Tcacao\\_genome\\_v1.1#tripal\\_analysis-downloads-box](http://www.cacaogenomedb.org/Tcacao_genome_v1.1#tripal_analysis-downloads-box); downloaded on 31<sup>st</sup> May 2015; 59,577 sequences; 23,720,084 residues). The amino acid sequence of BSA was added to the *Theobroma cacao* database.

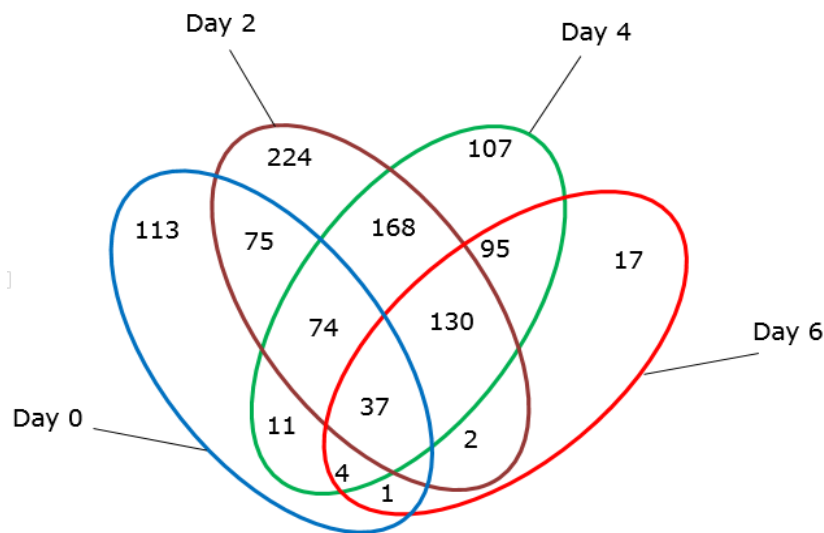
Searches were performed using parameter settings as previously published (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2018) with trypsin selected as enzyme for proteomic analysis while no enzyme was selected for peptidomic analysis, setting the false discovery rate (FDR) to 1% for all searches. For proteomic analysis only, searches were also carried out against a custom-made contaminants database, containing mainly trypsin and keratin-related proteins (70 sequences; 31,845 residues). Label-free protein quantitation was carried out using the average of the top 3 unique peptides for each protein. Protein intensities were normalised against BSA. Only proteins with a minimum of two unique peptides were quantified. For statistical analyses, JMP Pro 13.0 (SAS, Marlow, UK) and XLSTAT 2108.5 (Addinsoft, Paris, France) software were used.

## **3. Results**

Cocoa beans from the genotype IMC 67 were fermented in a Styrofoam box, and aliquots of samples taken at day 0, 2, 4, and 6 at the same day time. To evaluate how the peptide profiles changed during fermentation, UHPLC-ESI MS/MS analysis was performed on bean aliquots from the four different fermentation time points. Four preparative replicates were prepared for each fermentation time point, and each preparative replicate was analysed by one UHPLC-ESI MS/MS run. Only peptides detected in at least three preparative replicates of a fermentation time point were selected for comparative peptide analysis. With this requirement, a total of 1058 peptides obtained from fermentation were identified and quantified (see Supplementary Table 1) across all fermentation time points. The mean of the intensities for the preparative replicates of the same fermentation time point was calculated for each quantified peptide. The fold differences between the beans at different

fermentation time points are reported as the ratio of the highest mean versus the lowest mean for each quantified peptide.

The highest number of peptides were detected in the beans sampled at 2 and 4 days of fermentation, with 710 and 626 entries, respectively. A significantly lower number of peptides was found in the beans at day 6 (286 entries), while 315 peptides were detected in the day 0 beans. The graphical distribution of the detected peptides for all the fermentation time points is shown as a Venn diagram in Figure 1.



**Figure 1.** Venn diagram detailing the number of peptides identified for each time point.

Almost half of the detected peptides derived from vicilin (502 entries), while a total of 144 peptides were associated with a 21 kDa albumin. A significant number of peptides could also be associated with the following proteins: lipoxygenase (60 entries), RmLC cupin (40 entries), glyceraldehyde-3-phosphate dehydrogenase (29 entries) and peroxygenase (28 entries). The complete list of proteins associated with each peptide is provided in Supplementary Table 2. The list of peptides originating from vicilin and the 21 kDa albumin are provided in Supplementary Tables 3 and 4, respectively.

The vast majority of the peptides detected in the day 2, 4 or 6 beans were either not present or detected at a lower level in the day 0 beans. The number of peptides detected at a higher level (compared to the other fermentation points) at each fermentation time point, and the normalised intensities of the corresponding proteins measured by label-free bottom-up quantitative proteomic analysis are shown in Table 1.

**Table 1.** Number of endogenous peptides detected at a higher level (compared to the other fermentation points) at each fermentation time point with the normalised intensities of the corresponding proteins.

Accession	Description	Fermentation time point				Total number of peptides
		Day 0	Day 2	Day 4	Day 6	
Thecc1EG020665t1	Vicilin-A	8	359	101	34	502
		248.9	181.6	64.8	57.6	
Thecc1EG012658t1	21 kDa seed albumin	3	43	83	15	144
		153.1	241.3	151.7	69.8	
Thecc1EG026543t1	Lipoxygenase 1	28	1	22	9	60
		6.7	11.7	18.1	12.9	
Thecc1EG041085t1	RmlC-like cupins superfamily protein	2	29	9	0	40
		5.6	3.7	0.5	0.2	
Thecc1EG017080t2	Glyceraldehyde-3-phosphate dehydrogenase C2	1	21	7	0	29
		4.1	4.6	3.7	3.0	
Thecc1EG020975t1	Peroxygenase 2	6	2	17	3	28
		1.8	2.9	1.6	0.7	
Thecc1EG004867t1	Acidic endochitinase	1	15	3	0	19
		2.9	4.5	2.2	0.6	
Thecc1EG021809t1	HSP20-like chaperones superfamily protein	5	2	12	0	19
		0.4	0.5	0.3	0.2	
Thecc1EG026589t1	Eukaryotic aspartyl protease family protein	0	6	8	1	15
		4.2	6.3	6.5	6.3	
Thecc1EG042481t1	Late embryogenesis abundant protein B19.1A	4	2	4	1	11
		0.9	1.8	0.6	0.3	
Thecc1EG016949t1	Oleosin family protein	8	0	3	0	11
		1.4	2.9	3.5	3.4	
Thecc1EG030267t1	Saposin-like aspartyl protease family protein	0	5	3	2	10
		4.1	6.5	8.5	8.5	
	Others	83	26	45	16	170
	<b>Total number of peptides at a higher level at the given fermentation point</b>	150	509	315	84	1058


Proteins with less than 10 peptides have been labelled as "Others". The grey rows show the number of endogenous peptides detected at a higher level at each fermentation time point, while the rows below the number of peptides show the normalised intensities of the proteins from the bottom-up label-free quantitation analysis. Only the number of endogenous peptides are reported for proteins labelled as "Others".

As vicilin and the 21 kDa albumin were the proteins contributing the vast majority of peptides, the sequence coverage and cleavage sites for these two proteins were assessed, see Supplementary Figures 1-2. A total of 246 and 108 cleavage sites were localised in the sequences of vicilin and 21 kDa albumin, respectively. The peptides, generated from vicilin, covered 68% of the sequence of this protein, while a sequence coverage of 76% was observed for the 21 kDa albumin.

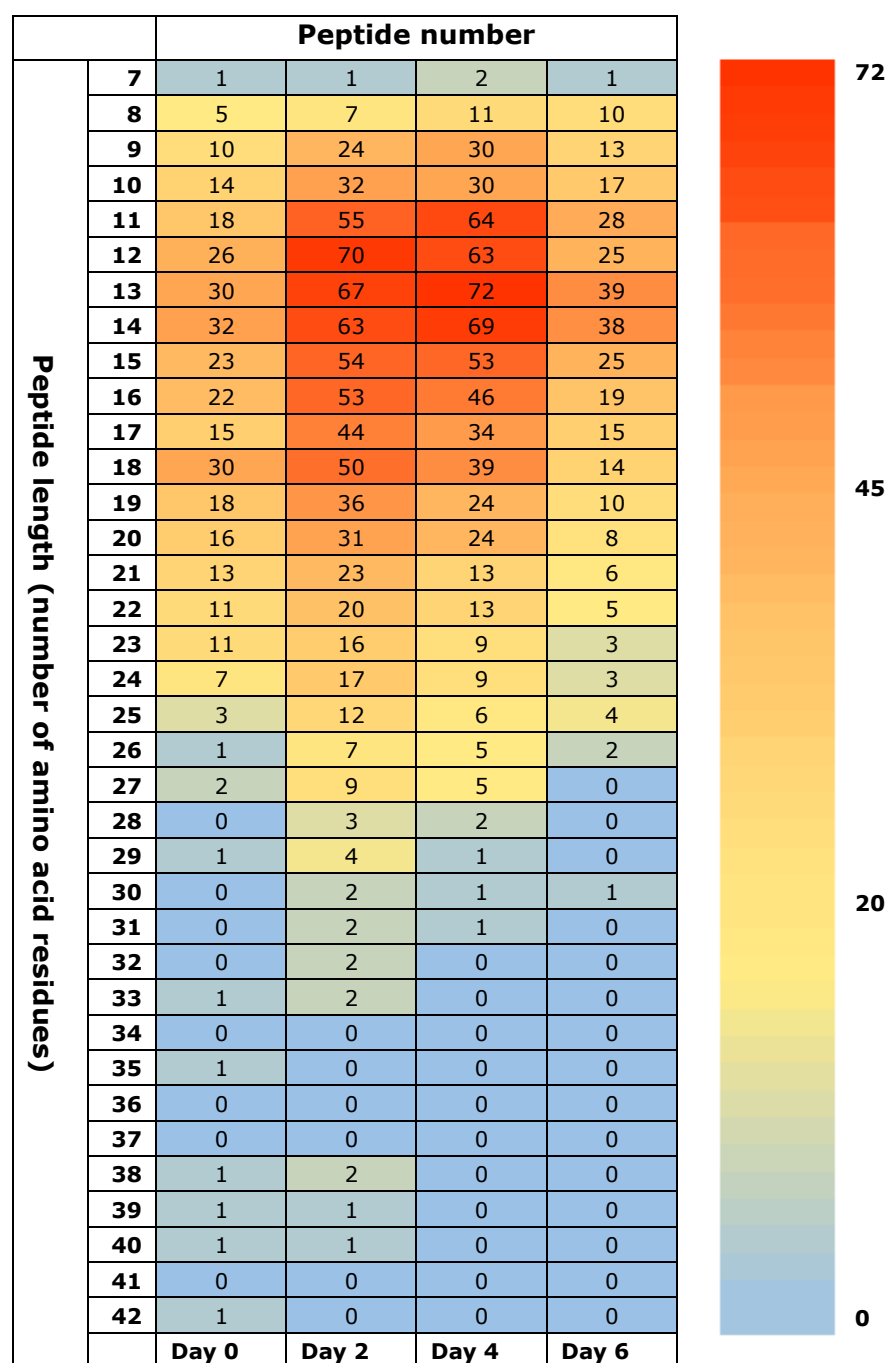
To assess how peptide length and intensity vary during fermentation, the sum of the MS ion signal intensities for all peptides with the same length (number of amino acid residues) for each fermentation time point is plotted in Table 2. In addition, the number of peptides with the same length (number of amino acid residues) for each fermentation time point is shown in Table 3.

**Table 2.** Sums of MS ion signal intensities for peptides with the same length (7-42 amino acid residues) detected at each fermentation time point.

		Sum of peptide ion signal intensity (arb. units)			
Peptide length (number of amino acid residues)	7	1.7E+06	3.4E+04	2.5E+05	1.5E+05
	8	3.5E+06	1.2E+07	1.9E+07	2.7E+07
	9	3.0E+06	6.8E+07	6.3E+07	4.9E+07
	10	9.0E+06	2.7E+07	2.4E+07	2.4E+07
	11	5.0E+06	1.1E+08	5.7E+07	1.9E+07
	12	1.0E+07	2.0E+08	8.3E+07	1.9E+07
	13	1.6E+07	1.8E+08	1.2E+08	4.0E+07
	14	3.0E+07	1.7E+08	7.3E+07	1.5E+07
	15	1.2E+07	9.9E+07	7.9E+07	2.8E+07
	16	1.0E+07	7.0E+07	1.1E+08	1.1E+07
	17	5.7E+07	1.9E+08	7.7E+07	2.3E+07
	18	7.4E+06	6.8E+07	4.3E+07	2.0E+07
	19	6.9E+07	1.5E+08	6.6E+07	1.3E+07
	20	3.6E+06	3.9E+07	1.7E+07	5.7E+06
	21	2.1E+06	3.2E+07	2.4E+07	1.0E+07
	22	2.0E+07	1.2E+08	5.6E+07	1.2E+07
	23	1.7E+07	5.6E+07	1.5E+07	2.6E+06
	24	7.8E+06	3.7E+07	1.7E+07	4.9E+06
	25	2.6E+05	2.3E+07	5.5E+06	1.1E+06
	26	2.3E+05	3.8E+07	2.4E+07	7.5E+06
	27	8.8E+05	7.8E+06	2.9E+06	0.0E+00
	28	0.0E+00	4.9E+06	2.2E+06	0.0E+00
	29	9.6E+04	2.4E+07	3.3E+06	0.0E+00
	30	0.0E+00	7.6E+05	2.1E+06	2.4E+05
	31	0.0E+00	2.1E+06	2.2E+05	0.0E+00
	32	0.0E+00	1.5E+06	0.0E+00	0.0E+00
	33	3.7E+05	7.9E+05	0.0E+00	0.0E+00
	34	0.0E+00	0.0E+00	0.0E+00	0.0E+00
	35	5.9E+04	0.0E+00	0.0E+00	0.0E+00
	36	0.0E+00	0.0E+00	0.0E+00	0.0E+00
	37	0.0E+00	0.0E+00	0.0E+00	0.0E+00
	38	4.4E+05	4.6E+06	0.0E+00	0.0E+00
	39	2.0E+05	2.8E+05	0.0E+00	0.0E+00
	40	2.0E+05	5.6E+05	0.0E+00	0.0E+00
	41	0.0E+00	0.0E+00	0.0E+00	0.0E+00
	42	7.4E+05	0.0E+00	0.0E+00	0.0E+00
		Day 0	Day 2	Day 4	Day 6



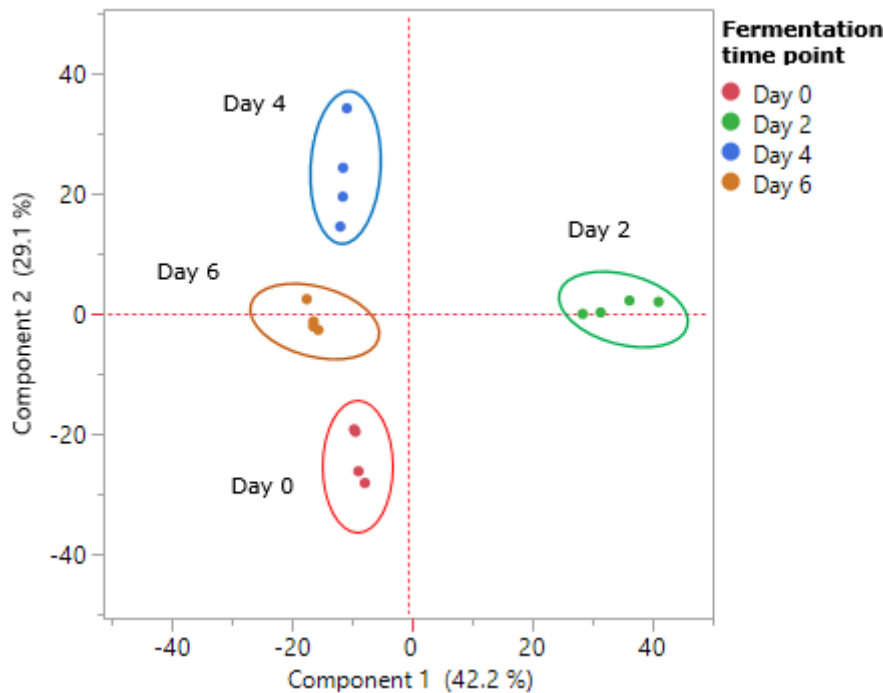
**Table 3.** Number of peptides with a length of 7-42 amino acid residues detected at each fermentation point.



The majority of the identified peptides showed an increase in intensity at fermentation time point day 2 and a significant reduction at the end of fermentation, with the exception of peptides with 8 residues which were more abundant at the end of fermentation. The highest MS ion signal intensities were observed for peptides with either 12-14, 17, or 19 residues at fermentation point day 2 (see Table 2) while the highest number of peptides was observed for peptides with a number of residues ranging from 11 to 15 at fermentation time points day 2 and 4 (see Table 3).

To understand whether peptides could be used as markers for the progress of fermentation, the most abundant peptides at each fermentation point (excluding day 0 beans) were selected. These peptides are listed in Supplementary Tables 5-7. Only peptides that met the following filters were included: highest abundance amongst the four fermentation time points with an RSD<20 for the peptide intensities among the preparative replicates of the same fermentation time point, a fold difference of >4 for the ratio between the highest and lowest mean intensity for each peptide across the different time points, and absence in the day 0 beans. With these filters, 192 potential markers were found for the fermentation time point at day 2, while 46 and 21 potential markers were obtained for the fermentation time points at day 4 and day 6, respectively.

To evaluate whether the data from the peptidomic analysis can be employed to graphically differentiate the four fermentation time points, PCA analysis was performed by loading the intensities of all detected peptides as variables and the fermentation time points as observations. In this case the data from all replicates were used. A clear separation of all fermentation time points can be observed on the PCA score plot based on the first two PCA components (see Figure 2).



**Figure 2.** PCA score plot of all (preparative) replicates using the peptide data listed in Supplementary Table 1. Preparative replicates of the same fermentation time point are displayed with the same colour.

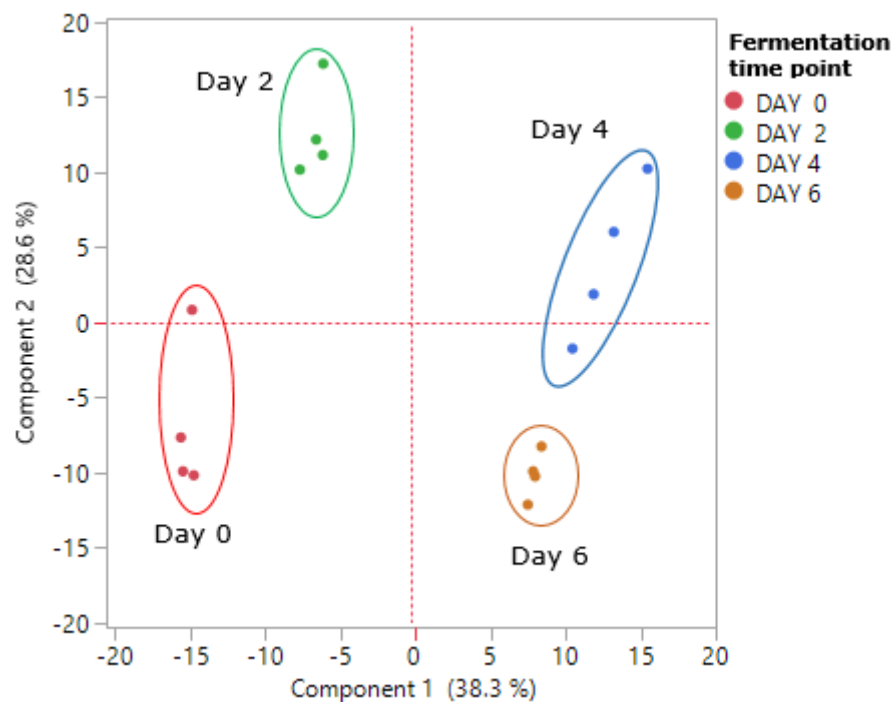
To understand the proteomic changes of cocoa beans in relation to the fermentation time, label-free quantitative bottom-up proteomic analysis was carried out on the same samples used to analyse the peptidome. The same number of preparative replicates were analysed, and proteins which were detected and quantified in at least three preparative replicates of a fermentation time point were evaluated. A total of 350 proteins were quantified when combining the entries of all fermentation time points. Protein intensities were normalised against BSA, which had been added as an internal standard to the samples prior to digestion, and the fold difference was calculated as previously described for the peptidome analysis. A list of all quantified proteins is provided in Supplementary Table 8.

The highest number of proteins were found in the beans of fermentation day 2 with 326 entries and day 0 with 325 entries. A lower number of proteins were detected in the beans at fermentation day 4 (268 entries) and day 6 (245 entries). The majority of the proteins (219 entries) were present at every fermentation time point. From the proteins detected at all fermentation time points, a total of 91 entries showed a fold difference of  $>2$  between at least one time point pair. A considerable number of proteins (131 entries) were detected in one or more but not all fermentation time points.

The total amount of proteins, based on the Bradford assay, showed a consistent decline over fermentation, starting at 10.0% (w:w) in the samples of the beans from day 0,

dropping to 8.2% after 2 days of fermentation, and 5.2% and 4.9% after 4 and 6 days of fermentation, respectively.

The normalised intensities of the proteins of all preparative replicates were loaded as variables in a PCA score plot, while the different fermentation time points were loaded as observations. This was to determine whether this classification would allow differentiating the fermentation time points based on the proteomic profile. As per the peptidome analysis, a clear separation of all fermentation time points can be observed, see Figure 3.



**Figure 3.** PCA score plot of all preparative replicates using the protein data listed in Supplementary Table 8. Preparative replicates of the same fermentation time point are displayed with the same colour.

Data supporting the results of this work are available in the PRIDE (Proteomics Identifications Database) partner repository at the European Bioinformatics Institute, PXD014434 (<http://www.ebi.ac.uk/pride/>).

#### 4. Discussion

The methodology employed in this study allowed the confident identification and quantitation of >1000 endogenous peptides, when combining the peptidomes of cocoa beans at different stages of fermentation, making it one of the largest sets of endogenous peptides ever reported for cocoa beans. Similar analyses had been previously undertaken (D'Souza, Grimbs, Grimbs, Behrends, & Corno, 2018; Neha Kumari, et al., 2018), identifying up to 900 peptides.



The detected peptidome varied substantially with fermentation time, with the number of identified peptides reaching a peak after 2 days of fermentation. A considerable drop in the number of identified peptides was observed in the beans at the final stage of fermentation. Only 37 peptides, were detected at all fermentation time points. The highest degree of overlap was found for the fermentation time points day 2 and day 4, with 168 common peptides. A total of 224 peptides were solely detected in the fermentation time point day 2.

It has been reported that the release of peptides from cocoa proteins is at its highest during the early stage of fermentation, and that the proteolytic activity is considerably reduced after three days (Lerceteau, Rogers, Petiard, & Crouzillat, 1999). This may explain why a higher number of peptides was detected in beans collected after 2 days of fermentation. However, a slightly higher number of peptides were detected in the beans at fermentation day 0 compared to the beans at the latest fermentation time point investigated (315 vs 286). These results are in contrast to the findings of Kumari *et al.* (Neha Kumari, et al., 2018) who suggested that well-fermented beans should have more than 300 peptides while beans with less than 100 peptides are to be considered under-fermented. However, in this study different cocoa varieties and fermentation procedures were employed. Also, cocoa pods were not immediately opened after harvest and as described in section 2.2, an aliquot of cocoa beans (day 0) was stored at -20° C overnight and freeze-dried for approximately 16 hours. Therefore, proteolysis by endogenous proteases might have occurred during this period, which could have led to the release of peptides even in beans that were not prepared for fermentation. However, endogenous proteases in cocoa beans are mainly activated during fermentation due to the generation of organic acids which lower the pH of the cocoa mass to levels which are optimal for these enzymes (Afoakwa, Paterson, Fowler, & Ryan, 2008; Schwan & Wheals, 2004).

Overall, the majority of the detected peptides originated from vicilin (502 entries), followed by 21 kDa albumin (144 entries), lipoxygenase (60 entries), RmLC cupin (40 entries), glyceraldehyde-3-phosphate dehydrogenase (29 entries) and a peroxygenase (28 entries). Vicilin and albumin are the most abundant cocoa bean proteins, and therefore it is reasonable to expect that their proteolytic products are well represented after fermentation. More than three times as many peptides originated from vicilin compared to 21 kDa albumin, reflecting its similarly greater number of amino acids. A high number of peptides generated from the degradation of these two proteins and a significant number of peptides related to lipoxygenase have already been reported by other authors (D'Souza, Grimbs, Grimbs, Behrends, & Corno, 2018).

The vast majority of vicilin peptides were present at a higher abundance at fermentation time point day 2, while almost 60% of the peptides derived from the 21 kDa albumin were

more abundant at fermentation time point day 4, see Table 1. This indicates that these two proteins are degraded at a different rate during fermentation, confirming previously published results (Caligiani, Marseglia, Prandi, Palla, & Sforza, 2016). Only a very small proportion of peptides derived from these two proteins were found at a higher level in the fermentation day 0 beans. Interestingly, no further degradation products of these peptides were found at the later stages of fermentation. In general, for these two proteins there is no clear observable trend of (smaller) subsequent degradation products being more abundant at later time points, indicating in many cases rapid peptide ladder degradation once a new protein fragment is produced.

The detected endogenous peptides from the 21 kDa albumin cover 73% of the sequence of this protein and were localised in specific zones spread throughout the sequence except for the initial 24 residues at the N-terminal, see Supplementary Figure 1. It has been reported that the initial 26 residues at the N-terminal of this protein constitute a signal peptide which is not present in the mature protein (Spencer & Hodge, 1991). However, the results of this work show that part of this peptide can be detected. Furthermore, peptides originating from cleavage sites within the initial 26 residues have been previously reported (D'Souza, Grimbs, Grimbs, Behrends, & Corno, 2018), indicating that the signal peptide may be still present in the mature protein. No peptides were detected for regions of the sequence localised at amino acid residues 109-116, 169-179, 214-221. The data allowed localising cleavage sites for aspartyl proteases and exopeptidases at both the N- and C-terminus. A peptide with the sequence SNADSKDDVVRVSTDVNIEF at position 89-108, possibly generated from the cleavage of an internal peptide bond by aspartyl protease activity, was further degraded into 9 smaller peptides by subsequent release of the C-terminal residue, likely due to the action of a carboxypeptidase, see Supplementary Table 4. N-terminal peptide cleavages were also observed such as by the degradation of the peptide TVWRLDNYDNSAGKW (position 120-134) into 6 smaller peptides as shown in Supplementary Table 4.

A similar sequence coverage was observed for the peptides released from vicilin (see Supplementary Figure 2). No peptides localised in the N-terminal region between 1-131 of the cocoa vicilin were identified, indicating that this N-terminus was not present at the time of fermentation or was not degraded during fermentation, as reported in previous studies (N. Kumari, et al., 2016; Marseglia, Sforza, Faccini, Bencivenni, Palla, & Caligiani, 2014; J. Voigt, Janek, Textoris-Taube, Niewianda, & Wostemeyer, 2016). Since no cocoa vicilin peptide has been previously identified from this region, it has been suggested that the annotation of vicilin at the N-terminus may not be correct (N. Kumari, et al., 2016). Examples of putative carboxypeptidase activity are found by the degradation of the peptide RSEEEEGQQRNNPYYFPKRSSFQTR at position 131-155 into 11 smaller peptides by

subsequent amino acid release at the C-terminus, see Supplementary Table 3. Fragmentation of the peptide RSEEEEGQQRNNPYYPFKR into 11 smaller peptides generated by subsequent cleavages of the N-terminal residue was a strong indication of aminopeptidase activity, see Supplementary Table 3.

One of the aims of this project was to evaluate whether peptides can be used as markers to assess the degree of cocoa bean fermentation. Thus, the peptides more abundant in the fermentation time points day 2, 4 and 6 are listed in Supplementary Tables 5, 6 and 7, respectively. However, in order to select strong marker candidates, only peptides that were not present in the day 0 beans, showing an RSD of <20 and fold difference of >4, were selected. The highest number of potential markers can be found for beans fermented for 2 days, while only 21 entries meet these criteria for fully fermented beans. Markers selected for the fermentation time point day 6 should be the most abundant peptides detected in beans that have undergone full fermentation. A higher abundance of markers selected for fermentation time points day 2 and day 4 could indicate that fermentation has not been achieved in full.

As there are quite a few potential markers showing differential abundance dependent on the fermentation time but no obvious peptide grouping that would display a fermentation time-dependent trend, PCA was deployed to evaluate its potential for distinguishing the four investigated fermentation time points (see Figure 2). Fermentation time points day 0, day 4 and day 6 were separated mainly by the first component, while fermentation time point day 2 was differentiated mainly based on the second component. Therefore, PCA as applied in this work is a useful tool to differentiate cocoa beans based on their fermentation stage.

Label-free quantitative bottom-up proteomic analysis was also employed in this study and based on the intensities of the most abundant unique tryptic digests for each protein. For the two main proteins, vicilin and 21 kDa albumin, a sharp decline of their normalised protein intensities obtained from this analysis was observed for the beans at the final fermentation stage compared to the day 0 beans (see Table 1). In agreement with the above-mentioned general lack of a trend amongst the entire set of endogenous peptides, these two major proteins followed a different degradation path during fermentation. Though the normalised intensity of vicilin showed a consistent decline as the fermentation progressed, the intensity for the 21 kDa seed albumin reached its highest level after 2 days of fermentation and started to decline after 4 days of fermentation to reach its lowest level in the fully fermented beans. Almost half of the endogenous peptides derived from lipoxygenase were present at a higher level in the day 0 beans, while the normalised intensity for this protein was higher in the later fermentation time points, suggesting that proteolytic activity on this protein during fermentation was only moderate. A similar trend

was observed for an oleosin protein. Apart from these and two aspartyl proteases, the normalised intensities of the proteins listed in Table 1 decreased over fermentation.

The total protein amount determined by the Bradford assay was highest in the day 0 beans, halved after 96 hours of fermentation and stabilized until the end of fermentation, indicating that protein degradation was significantly reduced after 96 hours of fermentation. The Bradford assay is based on the binding of the dye Coomassie Blue G250 to the arginyl and lysyl residues of proteins (Compton & Jones, 1985; Congdon, Muth, & Splittgerber, 1993). Therefore, a variation in response can be observed for different proteins. More importantly, the dye does not bind efficiently to peptides smaller than 3,000 Da, which could explain why the detected protein amount is reduced in the fermented beans.

Vicilin was the most abundant protein in the day 0 beans, followed by a 21 kDa albumin and 2S albumin. The intensities of these proteins varied over fermentation and a slightly higher abundance of the 21 kDa albumin with respect to vicilin was observed for the beans of the fermentation time point day 6. The ratio of the intensity of the 2S albumin versus both the 21 kDa albumin and vicilin showed a steady increase with fermentation time, see Supplementary Figure 3.

Unlike vicilin and the 21 kDa albumin, which released a high number of peptides during fermentation, the 2S albumin was not degraded during this process. This fact could be one of the reasons why the relative intensity of this protein increased over fermentation time. A lower amount of the highly abundant vicilin and 21 kDa albumin proteins may have also favoured the detection of the tryptic 2S albumin peptides, which would result in a higher intensity for this protein as well. The ratios of 2S albumin versus the 21 kDa albumin and vicilin could potentially be used to evaluate the degree of fermentation of cocoa beans.

Finally, the PCA score plot using the normalised protein intensities shows that the four fermentation time points were more spread out compared to the PCA score plot created by using the peptidomic data, thus potentially being more powerful for fermentation stage analysis.

## **5. Conclusion**

The data presented in this work allowed the identification and quantification of over 1000 endogenous peptides from cocoa beans at different fermentation stages, which makes it one of the largest sets of endogenous cocoa bean peptides reported to date. The number of identified cocoa bean peptides peaked at fermentation day 2 compared to day 0, 4, and 6. Most peptides originated from the abundant seed proteins vicilin and 21 kDa albumin. This data set enabled the evaluation of the peptidomic profile changes over fermentation,

and allowed the selection of potential markers that could be used to assess the degree of fermentation of cocoa beans. The abundance of the main cocoa proteins changed significantly over fermentation but followed different trends, providing an alternative to fermentation stage analysis of cocoa beans. PCA analysis of both the peptidomic and proteomic data allowed a clear separation of the beans at different fermentation stages. Thus, this methodology could be employed as a tool for creating larger databases of fermentation-relevant proteins and peptides and for an easy differentiation of beans at different fermentation stages.

## **6. Acknowledgements**

This work was financially supported by Mondelēz International.

## 523 7. References

- 524 Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and  
525 character in cocoa and chocolate: A critical review. *Critical Reviews in Food Science*  
526 *and Nutrition*, 48(9), 840-857.
- 527 Bradford, M. M. (1976). A Rapid and Sensitive Method for the Quantitation of Microgram  
528 Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical*  
529 *Biochemistry*, 72(1-2), 248-254.
- 530 Caligiani, A., Marseglia, A., Prandi, B., Palla, G., & Sforza, S. (2016). Influence of  
531 fermentation level and geographical origin on cocoa bean oligopeptide pattern. *Food*  
532 *Chemistry*, 211, 431-439.
- 533 Compton, S. J., & Jones, C. G. (1985). MECHANISM OF DYE RESPONSE AND  
534 INTERFERENCE IN THE BRADFORD PROTEIN ASSAY. *Analytical Biochemistry*, 151(2),  
535 369-374.
- 536 Congdon, R. W., Muth, G. W., & Splittgerber, A. G. (1993). THE BINDING INTERACTION  
537 OF COOMASSIE BLUE WITH PROTEINS. *Analytical Biochemistry*, 213(2), 407-413.
- 538 D'Souza, R. N., Grimbs, A., Grimbs, S., Behrends, B., & Corno, M. (2018). Degradation of  
539 cocoa proteins into oligopeptides during spontaneous fermentation of cocoa beans.  
540 *Food Research International*, 109, 516-516.
- 541 Kumari, N., Grimbs, A., D'Souza, R. N., Verma, S. K., Corno, M., Kuhnert, N., & Ullrich,  
542 M. S. (2018). Origin and varietal based proteomic and peptidomic fingerprinting of  
543 Theobroma cacao in non-fermented and fermented cocoa beans. *Food research*  
544 *international (Ottawa, Ont.)*, 111, 137-147.
- 545 Kumari, N., Kofi, K. J., Grimbs, S., D'Souza, R. N., Kuhnert, N., Vrancken, G., & Ullrich,  
546 M. S. (2016). Biochemical fate of vicilin storage protein during fermentation and  
547 drying of cocoa beans. *Food Research International*, 90, 53-65.
- 548 Lerceteau, E., Rogers, J., Petiard, V., & Crouzillat, D. (1999). Evolution of cacao bean  
549 proteins during fermentation: a study by two-dimensional electrophoresis. *Journal of*  
550 *the Science of Food and Agriculture*, 79(4), 619-625.
- 551 Marseglia, A., Sforza, S., Faccini, A., Bencivenni, M., Palla, G., & Caligiani, A. (2014).  
552 Extraction, identification and semi-quantification of oligopeptides in cocoa beans. *Food*  
553 *Research International*, 63, 382-389.
- 554 Schwan, R. F., & Wheals, A. E. (2004). The microbiology of cocoa fermentation and its  
555 role in chocolate quality. *Critical Reviews in Food Science and Nutrition*, 44(4), 205-  
556 221.
- 557 Scollo, E., Neville, D., Oruna-Concha, M. J., Trotin, M., & Cramer, R. (2018).  
558 Characterization of the Proteome of Theobroma cacao Beans by Nano-UHPLC-ESI  
559 MS/MS. *Proteomics*, 18(3-4).

560 Scollo, E., Neville, D. C. A., Oruna-Concha, M. J., Trotin, M., & Cramer, R. (2019). UHPLC-  
561 MS/MS analysis of cocoa bean proteomes from four different genotypes. *Food*  
562 *Chemistry*, 303, 8.

563 Spencer, M. E., & Hodge, R. (1991). Cloning and sequencing of the cDNA-encoding the  
564 major albumin of *Theobroma cacao*-Identification of the protein as a member of the  
565 Kunitz protease inhibitor family. *Planta*, 183(4), 528-535.

566 Voigt, J., Biehl, B., Heinrichs, H., Kamaruddin, S., Marsoner, G. G., & Hugi, A. (1994). In-  
567 vitro formation of cocoa-specific aroma precursors: aroma-related peptides generated  
568 from cocoa-seed proteins by co-operation of an aspartic endoprotease and a  
569 carboxypeptidase. *Food Chemistry*, 49(2), 173-180.

570 Voigt, J., Biehl, B., & Wazir, S. K. S. (1993). The major seed proteins of *Theobroma cacao*  
571 *L.* *Food Chemistry*, 47(2), 145-151.

572 Voigt, J., Janek, K., Textoris-Taube, K., Niewianda, A., & Wostemeyer, J. (2016). Partial  
573 purification and characterisation of the peptide precursors of the cocoa-specific aroma  
574 components. *Food Chemistry*, 192, 706-713.

575 Voigt, J., Textoris-Taube, K., & Woestemeyer, J. (2018). pH-Dependency of the proteolytic  
576 formation of cocoa- and nutty-specific aroma precursors. *Food Chemistry*, 255, 209-  
577 215.

578 Warren A., J., Bottcher, N. L., Asskamp, M., Bergounhou, A., Kumari, N., Ho, P.-W.,  
579 D'Souza, R. N., Nevoigt, E., & Ullrich, M. S. (2019). Forcing fermentation: Profiling  
580 proteins, peptides and polyphenols in lab-scale cocoa bean fermentation. *Food*  
581 *Chemistry*, 278, 786-794.

582